BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 4: e59070
doi: 10.3897/biss.4.59070

OPEN ACCESS

Conference Abstract

DINA—Development of open source and open services for natural history collections & research

Falko Glöckler‡, James Macklin§, David Peter Shorthouse§, Christian Bölling‡, Satpal Bilkhu§, Christian Gendreau§

‡ Museum für Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany
§ Agriculture and Agri-Food Canada, Ottawa, Canada

Corresponding author: Falko Glöckler (falko.gloeckler@mfn.berlin)

Received: 28 Sep 2020 | Published: 06 Oct 2020

Citation: Glöckler F, Macklin J, Shorthouse DP, Bölling C, Bilkhu S, Gendreau C (2020) DINA—Development of open source and open services for natural history collections & research. Biodiversity Information Science and Standards 4: e59070. https://doi.org/10.3897/biss.4.59070

Abstract

The DINA Consortium (DINA = "**DI**gital information system for **NA**tural history data", https://dina-project.net) is a framework for like-minded practitioners of natural history collections to collaborate on the development of distributed, open source software that empowers and sustains collections management. Target collections include zoology, botany, mycology, geology, paleontology, and living collections. The DINA software will also permit the compilation of biodiversity inventories and will robustly support both observation and molecular data.

The DINA Consortium focuses on an open source software philosophy and on community-driven open development. Contributors share their development resources and expertise for the benefit of all participants. The DINA System is explicitly designed as a loosely coupled set of web-enabled modules. At its core, this modular ecosystem includes strict guidelines for the structure of Web application programming interfaces (APIs), which guarantees the interoperability of all components (https://github.com/DINA-Web). Important to the DINA philosophy is that users (e.g., collection managers, curators) be actively engaged in an agile development process. This ensures that the product is pleasing for everyday use, includes efficient yet flexible workflows, and implements best practices in specimen data capture and management.

© Glöckler F et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

There are three options for developing a DINA module:

1. create a new module compliant with the specifications (Fig. 1),
2. modify an existing code-base to attain compliance (Fig. 2), or
3. wrap a compliant API around existing code that cannot be or may not be modified (e.g., infeasible, dependencies on other systems, closed code) (Fig. 3).

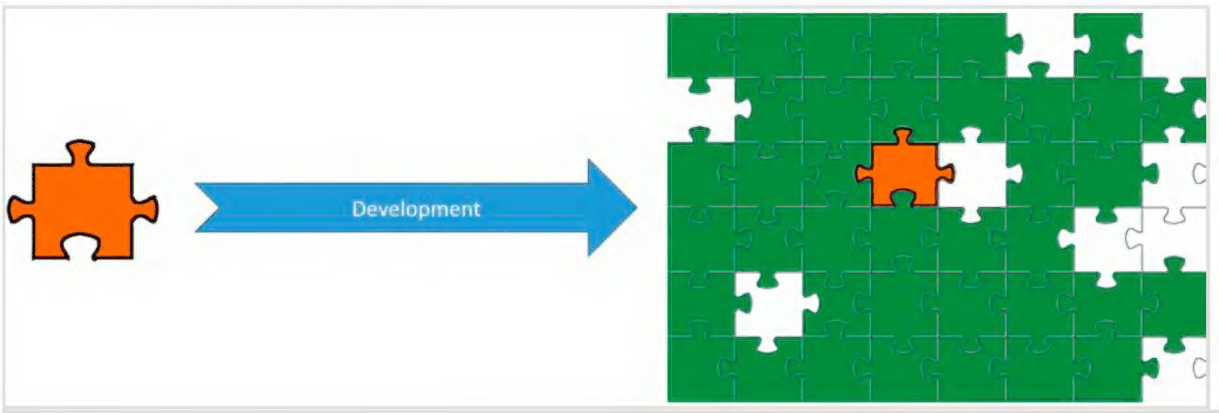

Figure 1.

First option to contribute to DINA developments: New code designed in a DINA compliant manner to allow for interoperability with the DINA system.

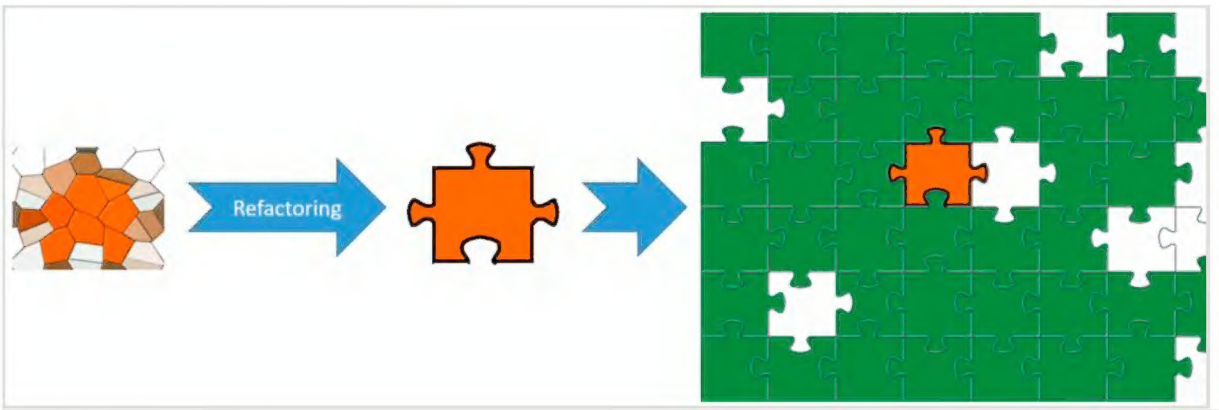

Figure 2.

Second option to contribute to DINA developments: Existing software can be refactored in a DINA compliant manner to allow for interoperability with the DINA system.

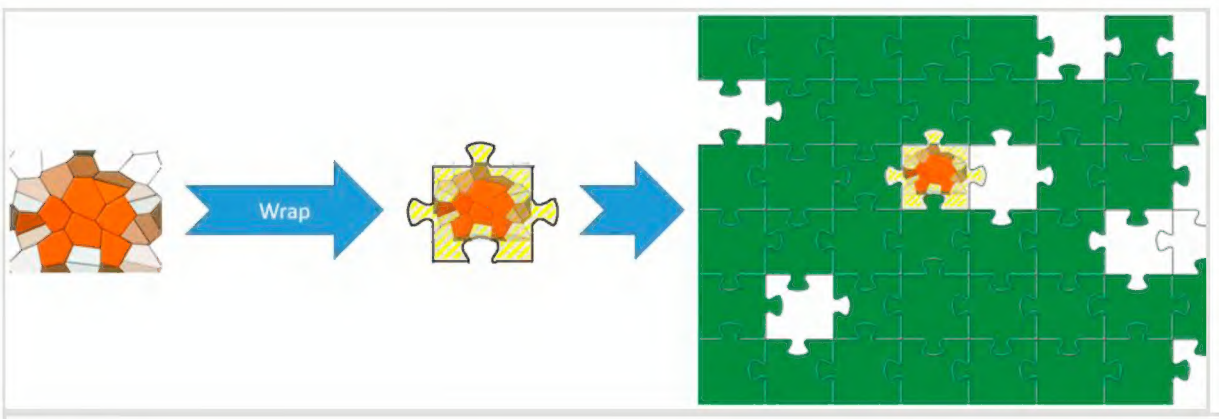

Figure 3.

Third option to contribute to DINA developments: Existing software that cannot be modified for several reasons could be wrapped by a DINA compliant API layer to allow for interoperability with the DINA system.

All three of these scenarios have been applied in the modules recently developed: a module for molecular data (SeqDB), modules for multimedia, documents and agents data and a service module for printing labels and reports:

The *SeqDB collection management and molecular tracking system* (Bilkhu et al. 2017) has evolved through two of these scenarios. Originally, the required architectural changes were going to be added into the codebase, but after some time, the development team recognised that the technical debt inherent in the project wasn't worth the effort of modification and refactoring. Instead a new codebase was created bringing forward the best parts of the system oriented around the molecular data model for Sanger Sequencing and Next Generation Sequencing (NGS) workflows.

In the case of the *Multimedia and Document Store* module and the *Agents* module, a brand new codebase was established whose technology choices were aligned with the DINA vision. These two modules have been created from fundamental use cases for collection management and digitization workflows and will continue to evolve as more modules come online and broaden their scope.

The *DINA Labels & Reporting* module is a generic service for transforming data in arbitrary printable layouts based on customizable templates. In order to use the module in combination with data managed in collection management software Specify (http://specifysoftware.org) for printing labels of collection objects, we wrapped the *Specify 7* API with a DINA-compliant API layer called the "*DINA Specify Broker*". This allows for using the easy-to-use web-based template engine within the DINA *Labels & Reports* module without changing Specify's codebase.

In our presentation we will explain the DINA development philosophy and will outline benefits for different stakeholders who directly or indirectly use collections data and related research data in their daily workflows. We will also highlight opportunities for joining the DINA Consortium and how to best engage with members of DINA who share their expertise in natural science, biodiversity informatics and geoinformatics.

Keywords

collection data management, research data, software development, collaboration, interoperability

Presenting author

Falko Glöckler

Presented at

TDWG 2020

Conflicts of interest

The DINA system is not intended to be a competitive product to other collection management systems. Instead, the DINA Consortium is pushing forward the approach of Open Source developments in a joint community-driven effort in order to maximize the usability and sustainability of the system beyond the availability of resources of individual institutions or vendors.

References

- Bilkhu S, El-Kayssi N, Poff M, Bushara A, Oh M, Giustizia J, Kandalaft I, Lowe C, Korol O, Sachs J, Newton K, Macklin J (2017) SeqDB: Biological Collection Management with Integrated DNA Sequence Tracking . Proceedings of TDWG 1 https://doi.org/10.3897/tdwgproceedings.1.20608